

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 02-083596

(43)Date of publication of application : 23.03.1990

(51)Int.CI. G10L 3/02
G10L 3/00

(21)Application number : 63-236911 (71)Applicant : MATSUSHITA ELECTRIC IND CO LTD

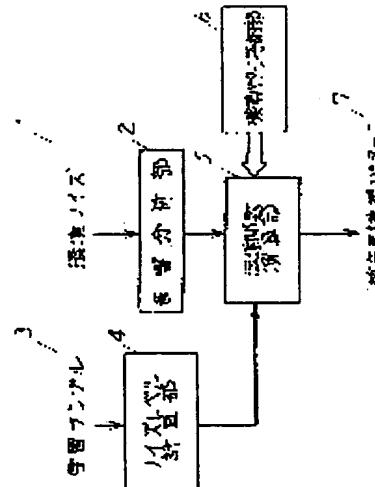
(22)Date of filing : 21.09.1988 (72)Inventor : MORII TOSHIYUKI
FUTAYADA KATSUYUKI

(54) SPEECH RECOGNIZING METHOD

(57)Abstract:

PURPOSE: To perform speech recognition matching an environmental noise and to improve a recognition rate by deriving a standard pattern from a artificial linear prediction coefficient representing a specific transfer function.

CONSTITUTION: The environmental noise 1 is inputted for a short time and an acoustic analyzing part 2 performs an acoustic analysis to find a linear prediction coefficient. A transfer function arithmetic part 5 adds the transfer function of a total-pole type composite model represented with the linear prediction coefficient for a phoneme standard pattern stored in a phoneme standard pattern storage part 6 and the transfer function of a total-pole type composite model represented with the linear prediction coefficient of the noise at noise level found by noise level learning and approximates the result as the total-pole type composite model to obtain a new transfer function. Then a new phoneme standard pattern is derived from the artificial linear prediction coefficient representing the new transfer function. This new phoneme standard pattern is used for the phoneme decision part of a speech recognition system to realize speech recognition matching the noise.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

⑯ 日本国特許庁 (JP)

⑪ 特許出願公開

⑯ 公開特許公報 (A)

平2-83596

⑯ Int. Cl. 5

G 10 L 3/02
3/00

識別記号

厅内整理番号

3 0 1 8842-5D
3 0 1 C 8842-5D

⑯ 公開 平成2年(1990)3月23日

審査請求 未請求 請求項の数 2 (全7頁)

⑯ 発明の名称 音声認識法

⑯ 特願 昭63-236911

⑯ 出願 昭63(1988)9月21日

⑯ 発明者 森井 利幸 神奈川県川崎市多摩区東三田3丁目10番1号 松下技研株式会社内

⑯ 発明者 二矢田 勝行 神奈川県川崎市多摩区東三田3丁目10番1号 松下技研株式会社内

⑯ 出願人 松下電器産業株式会社 大阪府門真市大字門真1006番地

⑯ 代理人 弁理士 粟野 重孝 外1名

明細書

1. 発明の名称

音声認識法

2. 特許請求の範囲

(1) 認識時の環境ノイズを音響分析して得られた線型予測係数で表現される全極型合成モデルの伝達関数と、標準パターンに対応する線型予測係数で表現される全極型合成モデルの伝達関数とを加え、その結果を全極型合成モデルとして近似することによって新たな伝達関数を得、その新たな伝達関数を表現する疑似線型予測係数から標準パターンを導くという、標準パターンのノイズ適合手段と、前記標準パターンのノイズ適合の、伝達関数の演算におけるノイズレベルを、認識時において集音する学習サンプルより導くノイズレベル学習手段とによって構成され、前記構成によってノイズ適合をなされた標準パターンをマッチングに用いることによって音声認識を行なうことを特徴とした音声認識法。

(2) 音声認識が音素を認識の基本単位としており、

標準パターンが線型予測係数から導かれるLPCケプストラム係数を特徴パラメータとする音素標準パターンであり、またマッチングが音素標準パターンを用いた音素マッチングであり、ノイズレベル学習手段がノイズレベルを、認識時において集音する学習サンプルの無音声区間と母音区間との平均パワーの比とするものであることを特徴とする請求項1記載の音声認識法。

3. 発明の詳細な説明

産業上の利用分野

本発明は、音声データを認識する装置に用いられる音声認識法に関するものである。

従来の技術

現在の音声認識システムの機能ブロック図は、第2図に示す通りである。まず、マイク入力された入力音声信号8に、LPC分析やフィルター分析などの音響分析をほどこし、特徴パラメータを含む音声情報を音響分析部9で抽出する。次に、セグメンテーション部10において子音セグメンテーションを行った後、音素判別部11において

音素標準パターン格納部12に格納された音素標準パターンとマッチングを行い、音素を判別して音素系列を作る。そして、音素系列作成部13において、音形規則格納部14に格納された音形規則と照し合せて修正をくわえて、最終的な音素系列を作成し、この音素系列と単語辞書格納部15に格納された単語辞書とのマッチングを単語マッチング部16で行って、類似度の一番大きいものを認識結果17とする。

ここで、音素判別部の音素マッチングに用いられる不特定話者・多数語用音素標準パターンは第3図に示す様に作成される。まず、防音室内でマイク入力された音声データ18をA/D変換部19においてA/D変換したものを多人数・多数語について収録して音声データベース20を作る。次に、音響分析部21で音響分析を行い、特徴パラメータを抽出する。一方、音響分析部21で得られるパワー情報をもとに、人が目視によってそれぞれのデータに対して音素のラベリング22を行って、ラベリングデータベース23を作る。

いて、指定された比で混合し、ノイズ付加音声データベース31を作成する。次に、音響分析部32で音響分析を行い、特徴パラメータを抽出する。次に、第3図において示された音素標準パターン作成において作成され、ラベリングデータベース格納部33に格納されたラベリングデータベースと特徴パラメータとを用いて、標準パターン作成部34において、LPCケブストラム係数を特徴パラメータとしたノイズ付加音素標準パターン35を作成する。

このノイズ付加音素標準パターンを音素判別に用いることによってノイズに適合した音声認識をおこなうことができるが、ノイズ付加音素標準パターン作成には膨大な時間と労力を必要とするので、認識時のノイズ学習では実現が困難であった。

発明が解決しようとする課題

音声認識システムの実用化に際しては、認識される音声が環境から受ける影響を低減する認識アルゴリズムが必要になる。特に、環境ノイズは認

して、標準パターン作成部24において、特徴パラメータとラベリングデータとを用いてLPCケブストラム係数を特徴パラメータとした音素標準パターン25を作成する。

この音声認識システムの実用化に際しては、認識される音声が環境から受ける影響を低減する認識アルゴリズムが必要になる。特に、環境ノイズは認識率の低下の最も大きい要因であり、音響分析におけるノイズ対策は必須である。従来の環境ノイズ対策は、その環境ノイズを付加した音素標準パターンを作成して音素認識に用いるという方法であったが、このノイズ付加音素標準パターンは第4図に示す様にして作成される。まず、あらかじめ音声認識を行なう環境のノイズ26を収録し、A/D変換部27においてA/D変換を行なつてノイズデータベース28を作成する。次に、第3図において示された音素標準パターン作成において作成され、音声データベース格納部29に格納されたノイズのないクリーンな音声データベースとノイズデータベースとをノイズ混合部30にお

認識率の低下の最も大きい要因であり、音響分析におけるノイズ対策は必須である。しかし、環境ノイズは、その環境や時間によって様々なパワーと周波数成分を持つので、それに対する適応が容易でない。また、LPCケブストラム係数を特徴パラメータとしているため、ノイズ成分の扱いが難しくなっている。そのため、現在の音声認識における音素標準パターンの環境ノイズ適合法としては、その環境ノイズを付加した音素標準パターンを作成して音素認識に用いるという方法が認識率向上に最も有効なものであった。

しかし、ノイズ付加音素標準パターンを作成するためには、その環境ノイズを収録し、それを音声データに付加したものを多人数・多数語について収集し、その音声データから音声データベースを作り、その音声データベースから音素標準パターンを作成するという大変労力と時間のかかるデータ処理を行わなくてはならない。また、いくつかの環境ノイズを付加した音素標準パターンを格納しておき、認識時にその中から最も適した音素

標準パターンを選択して音素認識に使用するという方法も考えられるが、ノイズのパワースペクトル上にピークがある場合などを含めると、それだけではすべての環境ノイズに対して対応することは出来ない。また、実環境で用いられる音声認識装置に組み込むということを考えると、認識時における環境ノイズ学習によって音素標準パターンのノイズ適合を行うことが望ましい。

本発明は、認識時における短時間の環境ノイズを音響分析して得られる特徴パラメータと、ノイズレベル学習によって得られるノイズレベルによって、音素標準パターンをその環境ノイズに適合するように変形し、その音素標準パターンを音素判別部における音素マッチングに使用することによって、環境ノイズに適合した音声認識を行うことによる認識率の向上を目的とするものである。

課題を解決するための手段

この課題を解決するために、本発明は線型予測係数で表現される全極型合成モデルの伝達関数の演算による音素標準パターンのノイズ適合を提案

て容易に得ることが出来る。その音素標準パターンを音素判別部における音素マッチングに使用することによって、環境ノイズに適合した音声認識を行うことが出来る。よって認識率の向上が実現される。

実施例

以下に本発明の実施例を図面を用いて詳細に説明する。第1図は、本発明の一実施例における音声認識システムを具体化する、線型予測係数で表現される全極型合成モデルの伝達関数の演算による音素標準パターンのノイズ適合の機能ブロック図である。

各ブロックの説明を以下に述べる。

まず認識時において、環境ノイズ1を短時間入力し、音響分析部2において音響分析を行ない、線型予測係数を求める。

次に、ノイズレベル学習のために、無音声区間と母音区間とから構成される学習サンプル3を発声・入力し、ノイズレベル計算部4において無音声区間と母音区間との平均パワーの比を計算する。

する。

まず認識時における環境ノイズを短時間収音し、そのノイズデータを音響分析して線型予測係数を求める。次に、ノイズレベル学習のために無音声区間と母音区間から構成される学習サンプルを発声・集音し、無音声区間と母音区間のパワーの比を求める。そして、音素標準パターンに対する線型予測係数で表現される全極型合成モデルの伝達関数と、求めたノイズの線型予測係数で表現される伝達関数とを、上記ノイズ学習によって求められたノイズレベルで加え、その結果を全極型合成モデルとして近似することによって新たな伝達関数を得、その新たな伝達関数を表現する疑似線型予測係数から導かれる音素標準パターンを音素判別部に用いることによって、環境ノイズに適合した音声認識が行なわれ、上記目的が達成される。

作用

本発明により、様々な性質を持つ環境ノイズに適合した音素標準パターンを、認識時における短時間のノイズ音響分析とノイズレベル学習によっ

てこの無音声区間と母音区間にはそれぞれノイズのパワーが含まれており、そのパワー比を疑似的にノイズ環境のS/N比とする。このときの計算は次の様に行われる。

$$\text{ノイズレベル} = P_s / P_n$$

P_s : 母音区間の平均パワー

P_n : 無音声区間の平均パワー

このノイズレベルの平方根を伝達関数演算部へ送る。

次に、伝達関数演算部5において、音素標準パターン格納部6に格納された音素標準パターンに対する線型予測係数で表現される全極型合成モデルの伝達関数とノイズの線型予測係数で表現される全極型合成モデルの伝達関数とを、上記ノイズレベル学習で求められたノイズレベルで加え、その結果を全極型合成モデルとして近似して新たな伝達関数を得、その新たな伝達関数を表現する疑似線型予測係数から新音素標準パターンを導く。そして、この新音素標準パターンを音声認識システムの音素判別部に用いることにより、ノイズに

適合した音声認識が実現される。

ここで、上記の伝達関数演算部における処理について詳細に述べる。

線型予測係数で表現される全極型合成モデルの伝達関数は下の様な全極型フィルターの形で書かれる。

$$\sigma / A(z) = \sigma / (1 + \sum_n a_n z^{-n})$$

σ : 利得係数

a_n : 線型予測係数

このときの σ は、自己相関法による線型予測分析の全2乗誤差に等しく、スペクトルの側から見ると、モデルのスペクトルエネルギーとデータのスペクトルエネルギーを整合させるための利得因子となっている。

したがって、入力スペクトル $X(e^{i\theta})$ とモデルスペクトルは次式の様な関係がある。

$$P_w(X(e^{i\theta})) \approx P_w(\sigma / A(z))$$

$P_w(\cdot)$: ある一定時間内における時系列のパワー

のことから、音声データにノイズが付加出来

σ_A : 標準パターンに対応する利得係数

σ_B : ノイズから求めた利得係数

$A(z)$: 標準パターンに対応する線型予測係数

$B(z)$: ノイズから求めた線型予測係数

ここで、ノイズレベル d は音声とノイズの混合比であるが、上記変換式における平方根内の式に示されるように、音素標準パターンに混合するノイズデータのパワーを正規化する必要がある。そのため、ノイズデータの平均パワーとしてはノイズの線型予測分析を行った全ノイズデータの平均パワーを、音素標準パターンのパワーとしては音素標準パターンを作成した音素を含む全単語の全音声区間の平均パワーを用いて次式を計算する。

P_A = 全音声区間の平均パワー

P_B = 全ノイズデータの平均パワー

とすると、①式は

$$\sigma_A / A(z) + \sqrt{P_A / P_B} \cdot \sigma_B / B(z)$$

を計算する。

伝達関数のたし算の仕方は次の様にして行なわれる。まず、有理式としてたし算を行う。次に、

る様に線型予測係数をノイズに適合させることが出来るものと考えられる。この発想のもとに、音素標準パターンのノイズ適合を行なう。

ノイズ付加音声データ作成は下式の様に行われている。

$$X(n) + d \cdot \sqrt{P_w(X(n)) / P_w(Y(n))} \cdot Y(n)$$

$X(n)$: 音声信号の時系列

$Y(n)$: ノイズ信号の時系列

d : 指定するノイズレベル

まず、ある区間について音声データのパワーを求め、指定されたノイズレベルにあらかじめノイズのパワーを変更して音声データに数値的に加える。

これと同じ手順で、上記ノイズレベル学習によって求められたノイズレベルに基づいて、音素標準パターンに対する線型予測係数で表現される合成モデルの伝達関数を変化させる。上記のノイズレベルを d とすると、次の様な伝達関数の変形を行なう。

$$\sigma_A / A(z) + d \cdot \sqrt{P_w(\sigma_A / A(z)) / P_w(\sigma_B / B(z))} \cdot \sigma_B / B(z) = ①$$

和の分数式の分母の式を分子の式で割って、分子が1になる様にする。このときの打切り次数は使用する音素標準パターンが表現されるLPCケプストラム係数の次数を使用する。(本実施例では、分母の余った分数式は切捨てとする。)そして、分母の定数項で分母と分子を割り、分母の定数項が1になるようにすることによって、この伝達関数を全極型合成モデルとして近似する。最後に、この有理式を全極型合成モデルとしたときの分母の係数を疑似線型予測係数とし、分子を疑似利得定数として、新たな、LPCケプストラム係数を特徴パラメータとした音素標準パターンを導く。

音素標準パターンを構成するLPCケプストラム係数の平均は前述の様にして導くが、共分散行列については、短時間のノイズ学習でこのノイズ適合を実現するため、ノイズの共分散を混入することが出来ない。しかし、ノイズの混入によって音声の分散が広がると考えられるので、簡易的に次の様な変形を行う。

$$\widehat{Cov}(i, j) = Cov(i, j) + (\widehat{\mu}(i) - \mu(i))(\widehat{\mu}(j) - \mu(j))$$

$\mu(i) \text{ Cov}(i, j)$: もとの音素標準パターンの平均と共分散。

$\bar{\mu}(i)$: 前述の様にして作成された平均。

ここで、本発明の効果を示す音素判別実験を試みる。ただし、ノイズレベル学習の精度についての評価は実際のノイズ環境下における音声認識システムによる音素認識率によって示される性質のものであるので、検証が難しい。しかし、ノイズレベルを、評価データとしてのノイズ付加音声データを作成した実際の S/N 比からずらせた場合の音素判別実験により、ノイズレベルに多少のずれ(±5 dB程度)があつても効果が得られることを検証しており、ノイズレベル学習の精度は十分であることが推察される。よって、本実施例において示す音素判別実験結果は、ノイズレベル学習を行わずに、そのかわりに、評価データとなるノイズ付加音声データ作成の S/N 比付近のノイズレベルで行った、実験結果である。評価条件は次の通りである。評価対象音素群は母音・鼻音と

語中子音(有声破裂音/M, N, *, B, D, R, Z/、無声破裂音/C, P, T, K/、摩擦音/Z, S, H/)である。評価対象話者は男女各10人(計20人)であり、特に語中子音については、OPENの評価をするために、一人一人の評価に用いる音素標準パターンをその評価対象話者を除いた19人で作成する。評価データとして用いるノイズ付加音声データは、母音・鼻音については「擬似HOTHノイズ(-6 dB/oct)付加音声データ」、語中子音については「展示会の実環境ノイズ付加音声データ」であり、両者のノイズ付加音声データ作成時の実際のノイズレベルは両方とも S/N 比で15dBである。実験は15dBから±20dBの範囲で5dB刻みに9種類のノイズレベルについて行なわれ、判別率の最も良いものを判別結果とする。また、本発明によるノイズ適合のノイズ分析に用いられるノイズの入力時間は10秒である。

上記の条件に基づいて、ノイズの入っていないクリーンな音素標準パターンに本発明のノイズ適合を適用することによって導かれた新たな音素標準パターンを用いて評価を行なう。

以下の表は、それぞれの評価音素群の判別結果を表したものである。

以下余白

表 判定結果

(単位 %)

	/a/	/o/	/u/	/i/	/e/	N	A.v.
クリーンによる	93.34	87.11	88.22	77.50	92.97	68.06	84.53
ノイズ付加	93.54	92.59	78.68	88.30	92.72	88.85	89.15
本発明による	93.05	90.91	82.12	85.60	93.64	79.30	87.44

(単位 %)

(A) 母音・鼻音

(単位 %)

クリーンによる	66.4	64.2	62.8	61.8	61.1	60.8
ノイズ付加	79.1	77.4	76.6	76.1	75.5	75.3
本発明による	75.7	72.3	70.9	69.8	69.1	68.8

(B) 語中子音(左から0, ±1, ±2, ±3, ±4, ±5)

(単位 %)

クリーンによる	66.6	67.5	64.5	64.1	63.5	63.3
ノイズ付加	86.5	85.7	83.8	82.8	81.9	81.0
本発明による	82.6	83.3	80.4	79.8	79.0	78.3

(C) 有声破裂音

(単位 %)

クリーンによる	66.6	67.5	64.5	64.1	63.5	63.3
ノイズ付加	86.5	85.7	83.8	82.8	81.9	81.0
本発明による	82.6	83.3	80.4	79.8	79.0	78.3

(D) 無声破裂音

(単位 %)

クリーンによる	86.3	87.0	86.5	86.3	85.3	83.7
ノイズ付加	91.3	92.8	92.7	93.1	93.0	92.4
本発明による	87.8	88.8	89.7	89.7	89.4	88.8

第1段目がノイズの入っていないクリーンな音素標準パターンで評価した結果である。第2段目がノイズ付加音声データから作成した音素標準パターンで評価した結果である。第3段目が本発明によるノイズ適合により得られた音素標準パターンで評価した結果である。また、母音・鼻音に関しては音素ごとの認識率を、語中子音に関しては基準フレームをふらせた場合の認識率を示す。ノイズ付加音素標準パターンの認識率にはおよばないものの、本発明によるノイズ適合によって識別率が改善されていることがわかる。1段目と2段目の識別率の差を100%とした時、本発明が改善した識別率の割合は、本発明のノイズ適合によって、母音・鼻音が63.0%、語中子音でも有声破裂音が73.2%～55.2%、無声破裂音が80.4%～84.7%、摩擦音が30.0%～58.6%改善され、短時間のノイズ学習でありながらも、大きい効果がある事を示している。

発明の効果

以上の様に本発明は、線型予測係数で表現され

る全極型合成モデルの伝達関数の演算により、様々な性質を持つ環境ノイズに適合した音素標準パターンを、認識時の短時間のノイズ音響分析とノイズレベル学習によって容易に得ることが出来、その音素標準パターンを音素判別部における音素マッチングに使用することによって、環境ノイズに適合した音声認識を行うことが出来、よって認識率を向上させることが出来、本発明の効果は大きい。

4. 図面の簡単な説明

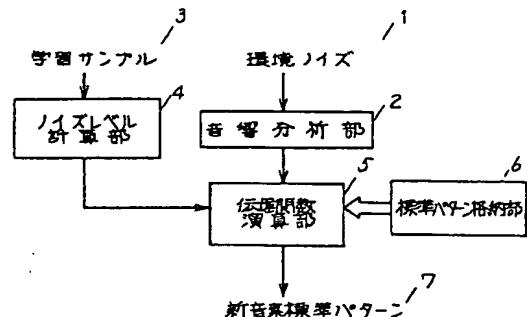
第1図は本発明の一実施例における音声認識システムを具体化する、線型予測係数で表現される合成モデルの伝達関数の演算による音素標準パターンのノイズ適合の機能ブロック図、第2図は従来の音声認識システムの機能ブロック図、第3図は従来の音素標準パターン作成システムの機能ブロック図、第4図は従来のノイズ付加音素標準パターン作成システムの機能ブロック図である。

1…環境ノイズ、2…音響分析部、3…学習サンプル、4…ノイズレベル計算部、5…伝達関数演算部

演算部、6…標準パターン格納部、7…新音素標準パターン。

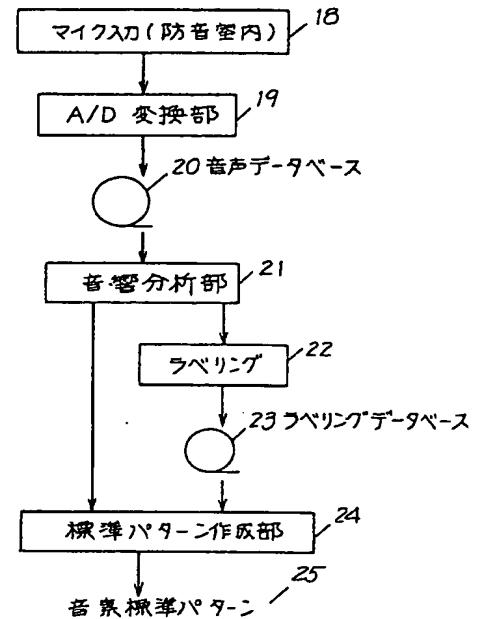
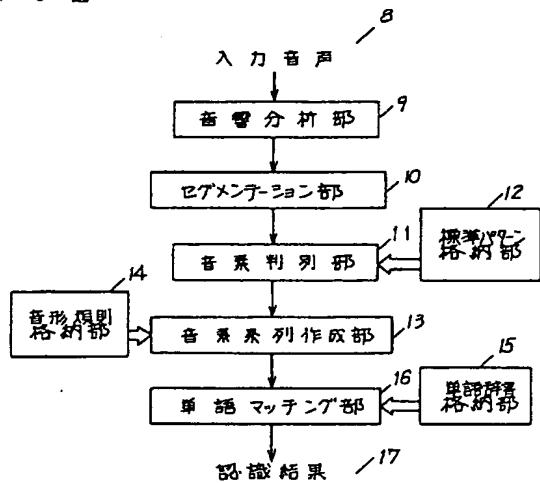
代理人の氏名 弁理士 篠 野 重 孝 ほか1名

第1図



第 3 図

第 2 図



第 4 図

